Retroviral G → A Hypermutation

Simon Wain-Hobson

Institut Pasteur, 28 rue du Dr. Roux, 75724 Paris, Cedex 15, France

A G → A hypermutant is a term given to retroviral proviruses bearing an inordinate number of G → A transitions, typical examples being given in Figure 1. Although two cases were unwittingly described for HIV-1 (Goodenow et al., 1989), the phenomenon was first described for spleen necrosis virus (SNV) (Pathak and Temin, 1990). More examples followed for HIV-1 (Vartanian et al., 1991; Vartanian et al., 1994), SIV (Johnson et al., 1991; Pelletier et al., 1995), HIV-2 (Gao et al., 1992), EIAV (Perry et al., 1992) and CAEV (Wain-Hobson et al., 1995). As the previous sentence shows, most have been described for members of the lentiviral group and much less frequently for other retroviruses. To some extent this undoubtedly reflects the current passion with anything lentiviral, meaning that there are huge data bases for the lentiviruses with respect to other retroviruses, particularly collections of PCR amplified material. However, it probably reflects a penchant of the lentiviral reverse transcriptase (Martinez et al., 1995). Note for the time being the absence of published examples for BIV, FIV and visna virus. For the record it is not an artefact confined to viral cultures as G → A hypermutants have been identified in uncultured PCR amplified material (Gao et al., 1992; Li et al., 1991; Pelletier et al., 1995).

```
AGCGGGGGAA TGATGGAGAA AGGAGAAATG
....A..A.. .A..A..A.. .AA.A.-..A
.......A.. AA..AA.A.. .AA.A....A


GCTGGGGAAA GATGGAGAGA GGAGAAATAA
.......... A..A....A. ..........
...AA.A... A...A.A.A. AA.A......
.......... A..AA...A. .A.A......
```

Figure 1. Some examples of hypermutated HIV gp120 env sequences.

Simple Criteria for Identifying G → A hypermutants

Nucleic acid sequence:

1) Monotony of G → A transitions with respect to the viral plus strand (Pathak and Temin, 1990; Vartanian et al., 1991). Only one recorded case of the phenomenon occurring during minus strand synthesis C → T hypermutation with respect to plus strand, (Pathak and Temin, 1990; Vartanian et al., 1991). Hypermutants may be occasionally accompanied by a few (<5%) other substitutions (Martinez et al., 1994).
2) All parts of the retroviral genome are vulnerable. There is one example of a provirus hypermutated throughout all 10 kb (Borman et al., 1995; Pelletier and Wain-Hobson, unpublished data).
3) Given that purine-purine transitions are the most frequent of all substitutions for lentiviruses, overall the number of G → A transitions per sequence should be ≥5 while the transition frequency should be >5% of the number of Gs. Up to 60% of Gs may be substituted (Wain-Hobson et al., 1995).
4) The distribution of substitutions may be confined to a very small region, say 50 bp (Delassus et al., 1991). Equally, they may be distributed in an erratic manner throughout the genome (Wain-Hobson et al., 1995). Finally hypermutation may be sustained throughout minus strand synthesis resulting in approximately 30% G substitution over a 10 kb proviral sequence (Borman et al., 1995; Pelletier and Wain-Hobson, unpublished data).

5) G → A transitions are associated with dinucleotide context declining in the order GpA > GpG > GpT ≥ GpC (Vartanian et al., 1991). Occasionally a few examples have GpG>GpA (Fitzgibbon et al., 1993; Vartanian et al., 1994). More complex sequence context rules have been proposed (Borman et al., 1995). This author is not convinced the latter will hold up.
6) Hypermutants may be accompanied by small deletions of 1–5 bases. Large deletions and small insertions (1–3 bases) are rarer (Pezo et al., 1996; Vartanian et al., 1991; Vartanian et al., 1994).

Protein sequence:

7) In phase stop codons resulting from tryptophan (TGG) to stop codons (TAA, TAG and TGA); see Figure 2.
8) Unusual proportions of certain amino acid substitutions, e.g., depletion of glycine (G → R, S, D, E and sometimes N or K), arginine (R → K, and less R → H or Q), serine (S → N), aspartic and glutamic acids (D → N and E → K respectively), valine (V → I), alanine (A → T), methionine (M → I) and cysteine (C → Y); see Figure 2.

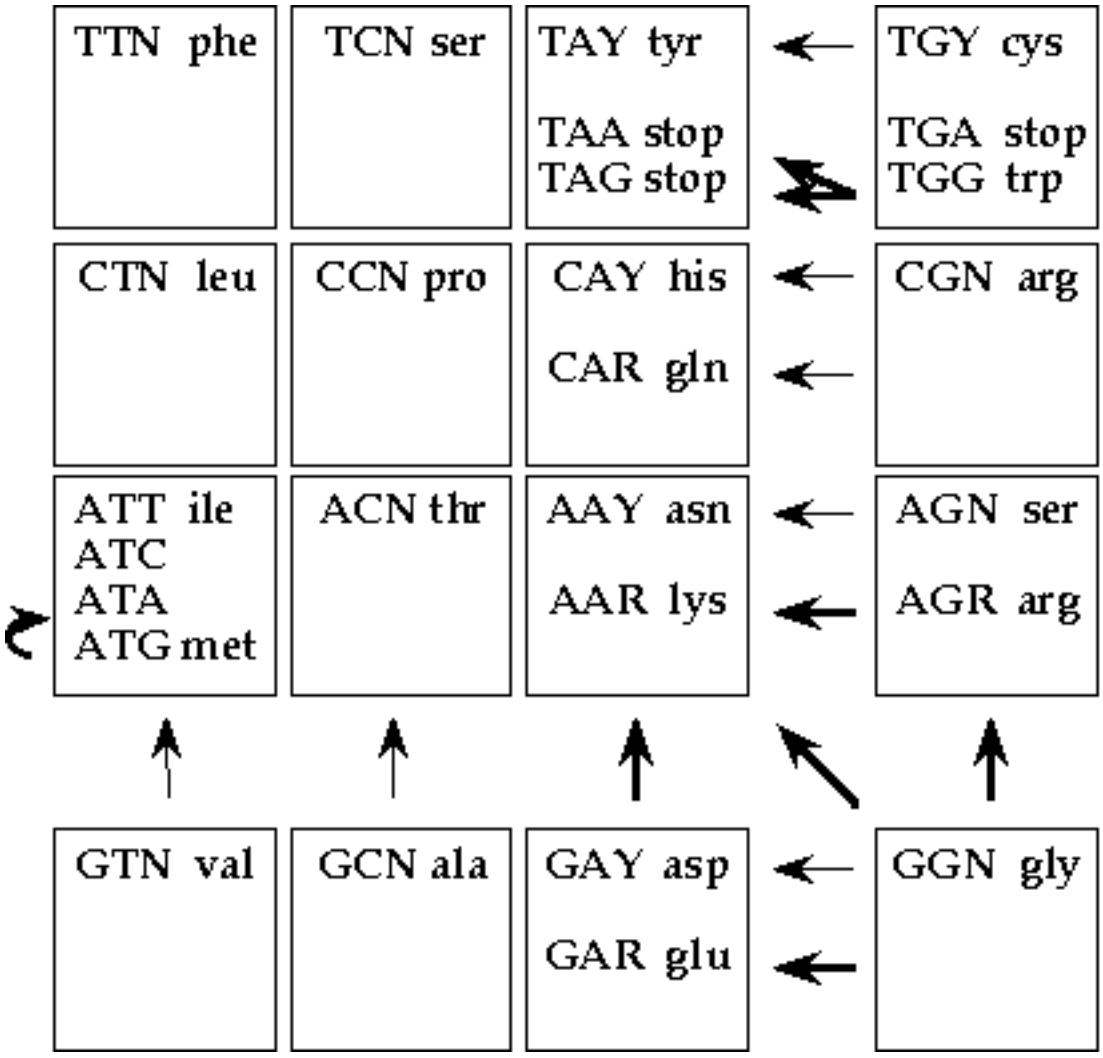

Figure 2. Amino acid substitutions resulting from G → A hypermutation. The intensity of the arrows indicates the relative frequency based generally on dinucleotide context.

Mechanism

G → A hypermutation results when minus strand DNA synthesis coincides with an increase in the intracellular [dTTP]/[dCTP] ratio.

Evidence

1) The two or so bases preceding the site of a misincorporation strongly influence the kinetics of misincorporation. Thus the strong dinucleotide context (Vartanian et al., 1991) indicates that it occurs during reverse transcription of plus stranded RNA into complementary DNA.
2) A sequence with a number of C → T transitions (with respect to the plus strand, which are but G → A transitions with respect to the minus strand) was identified along with many more G → A hypermutants (Vartanian et al., 1991). This indicated that the phenomenon could occur during both minus and plus DNA strand synthesis. The properties of RNA:DNA hybrids in vitro are more conducive to generation of rG:dT mismatches with respect to dG:dT mismatches in DNA:DNA duplexes (Sala et al., 1995).
3) G → A hypermutants may be produced in vivo by addition of deoxythymidine (dThd) to the culture supernatant (Vartanian et al., unpublished data). dThd is scavenged, and kinased up to dTTP which has a negative effect on ribonucleotide reductase reduction of CDP to dCDP. The net effect is to increase the [dTTP]/[dCTP] ratio.
4) G → A hypermutants may be simply reproduced in an in vitro reaction using RNA, purified RT and strongly biased [dTTP]/[dCTP] ratios (Martinez et al., 1994). They may also be made during endogenous strong stop synthesis (Martinez et al., 1994; Vartanian et al., unpublished data).
5) Many types of hypermutants can be produced in vitro by manipulation of the dNTP pools (Martinez et al., 1994; Vartanian et al., unpublished data). Thus the mutant spectrum is restricted by the intracellular milieu and much less the HIV-1 RT.
6) The extent of in vitro hypermutation was positively correlated to the magnitude of the [dTTP]/[dCTP] ratio (Martinez et al., 1994). The dinucleotide preference was reproduced at relatively low [dTTP]/[dCTP] ratio in the endogenous RT reaction (Vartanian et al., unpublished data).
7) HIV-1 RT is more sensitive in vitro to [dTTP]/[dCTP] imbalances than the Moloney murine leukemia virus (MoMLV) (Martinez et al., 1995). This helps explain the greater frequency of G → A hypermutants in lentiviral sequence data sets.
8) The clustering of rG:dT mismatches is partly explained by the observation that prior rG:dT mismatches increase the frequency of subsequent mismatches (Sala et al., 1995).

Additional negative arguments

1) Not associated with NTP imbalances during transcription. Given the pattern of reverse transcription the distribution of G → A transitions through contiguous U3 and R of the LTR proscribes a transcriptional origin (Vartanian et al., 1994).
2) Not related to excessive dUTP incorporation. Firstly hypermutants have been described for lentiviruses encoding a dUTPase (EIAV and CAEV, (Perry et al., 1992; Wain-Hobson et al., 1995)). Secondly for those without a dUTPase, such as HIV, the RT shows much greater discrimination against rG(template):dUTP as opposed to rGt:dTTP in vitro (Martinez et al., 1995). Finally, a dUTPase minus FIV mutant failed to generate G → A hypermutants (Lerner et al., 1995).
3) Not associated with a mutant polymerase. Hypermutants have been identified in a single cycle of reverse transcription with a "wild type" RT (Mansky, 1996; Pathak and Temin, 1990). Hypermutants can be produced in vitro using E. coli expressed RT derived from infectious molecular clones (Martinez et al., 1994).

Literature and Data Base

Overall hypermutants are rare, probably reflecting the rigours of purifying selection. There are probably fewer than 100 examples, many having been alluded to above. Some researchers relegate them

to the drawer as they are frequently defective. Others submit them to the data base without commenting on them in the conventional literature. Some examples go unperceived (Table 1).

Table 1 Some unsung or little known G → A hypermutants

Clone designation	Region of HIV genome*	Reference
ELI.03, TRA.18	env V1-V2	Goodenow et al 1989
clone 229	env V3 loop	LaRosa et al 1990
L3.14 (partially)	nef	Delassus et al 1991
YU, 1 clone	env V3-C4	Yu et al 1991
81wk:1–1, 1–3, 1–4, 1–11	SIVmac V1-V4	Overbaugh et al 1991
FOU 29.03.89 data set, 2 clones	p24 gag	Meyerhans et al 1991
FOI 21.05.90 data set, 1 clone		
MM152–12	SIVmac gag	Chen et al 1992
MM179–02, -10, -16, -22, -33		
316LSS3env	SIVmac V1-V5	Kodama et al 1993
D25/+24	pol	Najera et al 1995
A9 (Acc no. U28514)	SIVtan V3-V5	Mueller et al 1995
DH1 (partially)	nef	Huang et al 1995
BWB-11 (partially), -33	env	Monken et al 1995
CX-B (partially) patient Q23	V2 loop	Poss et al 1995
SP-2–203	nef	Michael et al 1995
D1.01	protease	Barrie et al 1996
pMCE10.86	LTR	Estable et al 1996
pMCE29.1		
HP93A1, A2, B1, B2, C1, C2	nef	Mariani et al 1996
HP95A1, A2		
HP83B1, B2 (perhaps)		

*HIV-1 unless otherwise stated. The table includes those sequences which passed into the literature with little or no comment. Other G → A hypermutants have been found for SNV (Pathak and Temin, 1990), HIV-1 (Fitzgibbon et al., 1993; Mansky, 1996; Vartanian et al., 1991; Vartanian et al., 1994), SIV (Johnson et al., 1991; Pelletier et al., 1995), HIV-2 (Gao et al., 1992), EIAV (Perry et al., 1992) and CAEV (Wain-Hobson et al., 1995). This list is probably not exhaustive, representing all those recognized by the author.

When hypermutation is not very intense or confined to a small region the molecular products cannot be assumed to be defective (Martinez et al., 1996). Thus, hypermutants should not be incorporated in phylogenetic trees, for they are generated in a single cycle of replication. There never were intermediates as is implied when constructing phylogeny. Doing so can lead to spurious conclusions or groupings (Mariani et al., 1996).

As mentioned above, one example was noted for SNV (murine C type retrovirus group) (Pathak and Temin, 1990). One case has been described for HTLV-1 being limited to 5 G → A transitions (4 GpA, 1 GpG) within a 110 bp stretch (i.e. 4.2% substitution frequency). In this context it is worth noting that during the endogenous cDNA reaction HTLV-1 hypermutants may be produced by using biased [dNTP]s (Vartanian et al., unpublished data).

And the Pararetroviruses?

This term covers the hepadnaviruses of mammals, reptiles and birds as well as the plant badnaviruses. Perhaps the most celebrated viruses of each group are human hepatitis B virus (HBV) and cauliflower mosaic virus (CaMV) respectively. Although the infectious form harbours a DNA genome, intracellular replication passes through reverse transcription of a plus strand RNA pre-genome. Are there any hypermutants? None have been published, although two such HBV genomes have been sequenced (Günther et al., unpublished data). Hypermutation was noted in the viral plus strand, while substitutions were erratically distributed throughout the genome. The overall substitution frequencies were 12% and 26%. Both genomes encoded in phase stop codons. The dinucleotide preference for transitions was GpA>GpG>GpC GpT. Such characteristics are typical of retroviral hypermutation, indicating that HBV replication is also vulnerable to intracellular [dTTP]/[dCTP] imbalances.

A → G Hypermutation

This represents an even rarer form of retroviral hypermutation, only a few examples being published to date. The initial example was identified within the U3 region of an avian leukosis virus (ALV) provirus (Felder et al., 1994), while more recent examples come from ALV (Hajjar and Linial, 1995) and SNV (Kim et al., 1996). They most certainly arose via post transcriptional deamination of adenine moieties in the RNA genome by double stranded adenosine deaminase. Such a phenomenon was first noted for measles virus genomes in cases of subacute sclerosing panencephalitis, leading to intervention in the post transcriptional modification of some cellular mRNAs. Deamination of adenine leaves inosine, which preferentially base pairs with cytosine, leading to A → G transitions. For review see Bass, 1995. Given the high A content of HIV, the extensive secondary structure in the region 1–600 and RRE, as well as the size of the HIV sequence data, it is perhaps surprising that no good example of A → G hypermutation exists. There has been discussion of an A/I change at position 27 of TAR (Blanchard et al., 1992; Klaver and Berkhout, 1994; Sharmeen et al., 1991).

Acknowledgements

Thanks to Drs. Eric Pelletier and Yoc Hoa Tran for computer analyses and Ben Berkout for email.

References

Barrie, K. A., Perez, E. E., Lamers, S. L., Farmerie, W. G., Dunn, B. M., Sleasman, J. W., and Goodenow, M. M. (1996). Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology* **219**, 407–416.

Bass, B. L. (1995). An I for editing. *Curr. Biol*. **5**, 598–600.

Blanchard, A. D., Powell, R., Braddock, M., Kingsman, A. J., and Kingsman, S. M. (1992). An adenosine at position 27 in the human immunodeficiency virus type 1 trans-activation response element is not critical for transcriptional activation by Tat. *J. Virol*. **66**, 6769–6772.

Borman, A. M., Quillent, C., Charneau, P., Kean, C. M., and Clavel, F. (1995). A highly defective HIV group O provirus: evidence for the role of local sequence determinants in hypermutation during negative strand DNA synthesis. *Virology* **208**, 601–609.

Chen, Z. C., Shen, L., Miller, M. D., Ghim, S. H., Hughes, A. L., and Letvin, N. L. (1992). Cytotoxic T lymphocytes do not appear to select for mutations in an immunodominant epitope of simian immunodeficiency virus gag. *J. Immunol*. **149**, 4060–4066.

Delassus, S., Cheynier, R., and Wain-Hobson, S. (1991). Evolution of human immunodeficiency virus type 1 nef and long terminal repeat sequences over 4 years in vivo and in vitro. *J. Virol*. **65**, 225–231.

Estable, M. C., Bell, B., Merzouki, A., Monyaner, J. S. G., O'Shaughnessy, M. V., and Sadowski, I. J. (1996). Human immunodeficiency virus type 1 long terminal repeat variants from 42 patients

representing all stages of infection display a wide range of polymorphism and transcription activity. *J. Virol*. **70**, 4053–4062.

Felder, M. P., Laugier, D., Yatsula, B., Dezélée, P., Calothy, G., and Marx, M. (1994). Functional and biological properties of an avian variant long terminal repeat containing multiple A to G conversions in the U3 sequence. *J. Virol*. **68**, 4759–4767.

Fitzgibbon, J. E., Mazar, S., and Dubin, D. T. (1993). A new type of G → A hypermutation affecting human immunodeficiency virus. *AIDS Res. Hum. Retroviruses* **9**, 833–838.

Gao, F., Yue, L., White, A. T., Pappas, P. G., Barchue, J., Hanson, A. P., Greene, B. M., Sharp, P. M., Shaw, G. M., and Hahn, B. H. (1992). Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. *Nature* **358**, 495–499.

Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J., and Wain-Hobson, S. (1989). HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. *J. Acquir. Immune Def. Syndr*. **2**, 344–352.

Günther, S., Sommer, G., Plikat, U., Iwanska, A., Wain-Hobson, S., Will, H., and Meyerhans, A. (unpublished data).

Hajjar, A. M., and Linial, M. L. (1995). Modification of retroviral RNA by double-stranded RNA adenosine deaminase. *J. Virol*. **69**, 5878–5882.

Huang, Y., Zhang, L., and Ho, D. D. (1995). Characterization of nef sequences from long-term survivors of human immunodeficiency virus type 1 infection. *J. Virol*. **69**, 93–100.

Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S., and Hirsch, V. M. (1991). The genetic fate of moleculary cloned simian immunodeficiency virus in experimentally infected macaques. *Virology* **185**, 217–228.

Kim, T., Mudry Jr., R. A., Rexrode II, C. A., and Pathak, V. K. (1996). Retroviral mutation rates and A-to-G hupermutations during different stages of retroviral replication. *J. Virol*. **70**, 7594–7602.

Klaver, B., and Berkhout, B. (1994). Premature strand transfer by the HIV-1 reverse transcriptase during strong-stop DNA synthesis. *Nucleic Acids Res*. **22**, 137–144.

Kodama, T., Mori, K., Kawahara, T., Ringler, D. J., and Desrosiers, R. C. (1993). Analysis of simian immunodeficiency virus sequence variation in tissues of rhesus macaques with simian AIDS. *J. Virol*. 67, 6522–6534.

LaRosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, R. N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Matthews, T. J., Emini, E., and Putney, S. D. (1990). Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science* **249**, 932–935.

Lerner, D. L., Wagaman, P. C., Phillips, T. R., Prospero-Garcia, O., Henriksen, S. J., Fox, H. S., Bloom, F. E., and Elder, J. H. (1995). Increased mutation frequency of feline immunodeficiency virus lacking functional deoxyuridine-triphosphatase. *Proc. Natl. Acad. Sci. USA*. **92**, 7480–7484.

Li, Y., Kappes, J. C., Conway, J. A., Price, R. W., Shaw, G. M., and Hahn, B. H. (1991). Molecular characterization of human immunodeficiency virus type 1 cloned directly from unclutured human brain tissue: identification of replication-competent and -defective viral genomes. *J. Virol*. **65**, 3973–3985.

Mansky, L. M. (1996). The mutation rate of human immunodeficiency virus type 1 is influenced by the vpr gene. *Virology* **222**, 391–400.

Mariani, R., Kirchhoff, F., Greenough, T. C., Sullivan, J. L., Desrosiers, R. C., and Skowronski, J. (1996). High frequency of defective nef alleles in a long-term survivor with non-progressive human immunodeficiency virus type 1 infection. *J. Virol*. **70**, 7752–7764.

Martinez, M. A., Pezo, V., Marlière, P., and Wain-Hobson, S. (1996). Exploring the functional robustness of an enzyme by in vitro protein evolution. *EMBO J*. **15**, 1203–1210.

Martinez, M. A., Sala, M., Vartanian, J. P., and Wain-Hobson, S. (1995). Reverse transcriptase and substrate dependence of the RNA hypermutagenesis reaction. *Nucleic Acids Res*. **14**, 2573–2578.

Martinez, M. A., Vartanian, J. P., and Wain-Hobson, S. (1994). Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc. Natl. Acad. Sci. USA*. **91**, 11787–11791.

Meyerhans, A., Dadaglio, G., Vartanian, J. P., Langlade-Demoyen, P., Frank, R., Asjo, B., Plata, F., and Wain-Hobson, S. (1991). In vivo persistence of a HIV-1-encoded HLA-B27-restricted cytotoxic T lymphocyte epitope despite specific in vitro reactivity. *Eur. J. Immunol*. **21**, 2637–2640.

Michael, N. L., Chang, G., D'Arcy, L. A., Tseng, C. J., Birx, D. L., and Sheppard, H. W. (1995). Functional characterization of human immunodeficiency virus type 1 nef genes in patients with divergent rates of disease progression. *J. Virol*. **69**, 6758–6769.

Monken, C. E., Wu, B., and Srinivasan, A. (1995). High resolution analysis of HIV-1 quasispecies in the brain. *AIDS*. **9**, 345–349.

Mueller-Trutwin, M. C., Corbet, S., Dias-Tavares, M., Herve, V. M. A., Nerrienet, E., Georges-Courbot, M. C., Saurin, W., Sonigo, P., and Barré-Sinoussi, F. (1995). HIV sequence data base.

Najera, I., Holguin, A., Quinones-Mateu, M. E., Munoz-Fernandez, M. A., Najera, R., Lopez-Galindez, C., and Domingo, E. (1995). pol gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. *J. Virol*. **69**, 23–91.

Overbaugh, J., Rudensey, L. M., Papenhausen, M. D., Benveniste, R., and Morton, W. R. (1991). Variation in simian immunodeficiency virus env is confined to V1 and V4 during progression to simian AIDS. *J. Virol*. **35**, 7025–7031.

Pathak, V. K., and Temin, H. M. (1990). Broad spectrum of in vitro forward mutations, hypermutations, and mutational hotspots in a retrovital shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl. Acad. Sci. USA*. **87**, 6019–6023.

Pelletier, E., Saurin, W., Cheynier, R., Letvin, N. L., and Wain-Hobson, S. (1995). The tempo and mode of SIV quasispecies development in vivo calls for massive viral replication and clearance. *Virology* **208**, 644–652.

Pelletier, E., and Wain-Hobson, S. (unpublished data).

Perry, S. T., Flaherty, M. T., Kelley, M. J., Clabough, D. L., Tronick, S. R., Coggins, L., Whetter, L., Lengel, C. R., and Fuller, F. (1992). The surface envelope protein gene region of equine infectious anemia virus is not an important determinant of tropism in vitro. *J. Virol*. **66**, 4085–4097.

Pezo, V., Martinez, M. A., and Wain-Hobson, S. (1996). Fate of direct and inverted repeats in the RNA hypermutagenesis reaction. *Nucleic Acids Res*. **24**, 253–256.

Poss, M., Martin, H. L., Kreiss, J. K., Granville, L., Chohan, B., Nyange, P., Mandaliya, K., and Overbaugh, J. (1995). Diversity in Virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. *J. Virol*. **69**, 8118–8122.

Sala, M., Wain-Hobson, S., and Schaeffer, F. (1995). HIV-1 reverse transcriptase tG:T mispair formation on RNA and DNA templates with mismatched primers: a kinetic and thermodynamic study. *EMBO J*. **14**, 4622–4627.

Sharmeen, L., Bass, B., Weintraub, H., and Groudine, M. (1991). Tat-dependent adenosine-to-inosine modification of wild-type transcativation response RNA. *Proc. Natl. Acad. Sci. USA*. **88**, 8096–8100.

Vartanian, J. P., Meyerhans, A., Åsjö, B., and Wain-Hobson, S. (1991). Selection, recombination and G → A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol*. **65**, 1779–1788.

Vartanian, J. P., Meyerhans, A., Sala, M., and Wain-Hobson, S. (1994). G → A hypermutation of the HIV-1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc. Natl. Acad. Sci. USA*. **91**, 3092–3096.

Vartanian, J. P., Plikat, U., Henry, M., Mahieux, R., Guillemot, L., Barré-Sinoussi, F., Meyerhans, A., and Wain-Hobson, S. (unpublished data).

Wain-Hobson, S., Sonigo, P., Guyader, M., Gazit, A., and Henry, M. (1995). Erratic G → A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology* **209**, 297–303.